

What is Claimed:

1. A method for specifying a pronunciation of a word comprising:  
receiving a written version of the word defined by a series of characters;  
separating the written version of the word into the series of characters;  
and  
generating symbols that define a pronunciation of the word based solely  
on the series of characters.
2. The method of claim 1, wherein receiving a written version of the  
word includes:  
receiving the written version of the word from a user.
3. The method of claim 1, wherein receiving a written version of the  
word includes:  
receiving the written version of the word from a program that automatically  
scans a network.
4. The method of claim 1, wherein the generated symbols have a  
one-to-one correspondence with the series of characters.
5. The method of claim 1, wherein the generated symbols correspond  
to predetermined character groupings from the series of characters.

6. The method of claim 5, wherein the predetermined character groupings are determined based on a statistical analysis of a language.
7. The method of claim 6, wherein the statistical analysis is based on frequency of occurrence of the words in the language.
8. The method of claim 1, further comprising:  
classifying the word into one of a predetermined plurality of classifications;  
and  
generating the symbols based on the classification of the word.
9. The method of claim 8, wherein the classifications are based on word affixes.
10. A speech recognition system comprising:  
speech recognition models configured to convert audio containing speech into a transcription of the speech;  
a system dictionary used to train the speech recognition models by providing symbols that define pronunciations of words to the speech recognition models; and  
a dictionary creation component configured to generate the symbols for the system dictionary, the symbols being based on written characters of the words.

11. The system of claim 10, wherein the dictionary creation component receives the words from a program that automatically scans a network for the words.

12. The system of claim 10, wherein the generated symbols have a one-to-one correspondence with a sequence of the written characters of the words.

13. The system of claim 10, wherein the generated symbols correspond to predetermined character groupings in a sequence of the written characters of the words.

14. The system of claim 13, wherein the predetermined character groupings are determined based on a statistical analysis of a language.

15. The system of claim 14, wherein the statistical analysis is based on frequency of occurrence of the words in the language.

16. The system of claim 10, wherein the dictionary creation component classifies each of the words into one of a predetermined plurality of classifications and generates the symbols based on the classifications.

17. A method comprising:

configuring a dictionary creation component to generate symbols that represent pronunciations of words in a target language, the symbols being generated based solely on written representations of the words and the configuring being performed based on the target language;

providing the dictionary creation component with written words; and

receiving the symbols that represent pronunciations of the written words from the dictionary creation component.

18. The method of claim 17, wherein the generated symbols have a one-to-one correspondence with a series of characters that define the written representations of the words.

19. The method of claim 17, wherein the generated symbols correspond to predetermined character groupings from a series of characters that define the written representations of the words.

20. The method of claim 19, wherein the predetermined character groupings are determined based on a statistical analysis of the target language.

21. The method of claim 20, wherein the statistical analysis is based on frequency of occurrence of the words in the target language.

22. The method of claim 17, further comprising:  
classifying the words into one of a predetermined plurality of  
classifications; and  
generating the symbols based on the classifications of the words.
23. The method of claim 22, wherein the classifications are based on  
word affixes.
24. A device comprising:  
means for receiving a written version of a word defined by a series of  
characters;  
means for separating the written version of the word into the series of  
characters; and  
means for generating symbols that define a pronunciation of the word  
based on the series of characters.
25. The device of claim 24, wherein the generated symbols have a  
one-to-one correspondence with the series of characters.
26. The device of claim 24, wherein the generated symbols correspond  
to predetermined character groupings from the series of characters.

27. The device of claim 26, wherein the predetermined character groupings are determined based on a statistical analysis of a language.

28. The device of claim 27, wherein the statistical analysis is based on frequency of occurrence of the words in the language.

29. The device of claim 24, further comprising:  
means for classifying the word into one of a predetermined plurality of classifications; and  
means for generating the symbols based on the classification of the word.

6

30. A computer-readable medium containing programming instructions for execution by a processor, the computer-readable medium comprising:

instructions for receiving a written version of a word defined by a series of characters;

instructions for separating the written version of the word into the series of characters; and

instructions for generating symbols that define a pronunciation of the word based solely on the series of characters.